

IN THE SPECIFICATION

Please amend paragraphs 80, 81, 84, 87, 92 and 93 of the specification as follows:

[0080] There are many properties of image appearance that one could use as data streams from which one could learn appearance models for tracking and object search. Examples include local color statistics, multiscale filter responses, and localized edge fragments. In this work, the data streams were derived from responses of a steerable filter pyramid is applied (i.e., based on the G_2 and H_2 filters; see W. Freeman and E. H. Adelson, "The Design and Use of Steerable Filters", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:891-906, 1991, ~~incorporated herein by reference~~). Steerable pyramids provide a description of the image at different scales and orientations that is useful for coarse-to-fine differential motion estimation, and for isolating stability at different scales and at different spatial locations, and different image orientations. Here G_2 and H_2 filters are used at two scales, tuned to wavelengths of eight and sixteen pixels (subsampling by factors of two and four), with four orientations at each scale.

[0081] From the filter outputs, the present inventors chose to maintain a representation of the phase structure as the appearance model. This provides a natural degree of amplitude and illumination independence, and it provides the fidelity for accurate image alignment afforded by phase-based methods (see, for example, D.J. Fleet and A.D. Jepson, "Stability of Phase Information", *IEEE Transactions on PAMI*, 15(12):1253-1268, 1993, ~~incorporated herein by~~

reference). Phase responses associated with small filter amplitudes, or those deemed unstable according to the techniques described in the above-cited paper were treated as outliers.

[0084] The motion is represented in terms of frame-to-frame parameterized image warps. In particular, given the warp parameters \mathbf{c}_t , a pixel \mathbf{x} at frame $t - 1$ corresponds to the image location $\mathbf{x}_t = \mathbf{w}(\mathbf{x}; \mathbf{c}_t)$ at time t , where $\mathbf{w}(\mathbf{x}; \mathbf{c}_t)$ is the warp function. Similarity transforms are used here, so $\mathbf{c}_t = (\mathbf{u}_t, \theta_t, \rho_t)$ is a 4-vector describing translation, rotation, and scale changes, respectively. Translations are specified in pixels, rotations in radians, and the scale parameter denotes a multiplicative factor, so $\bar{\eta} \equiv (0, 0, 0, 1)$ is the identity warp. By way of tracking, the target neighborhood is convected (i.e. warped) forward at each frame by the motion parameters. That is, given the parameter vector \mathbf{c}_t , \mathcal{N}_t is just the elliptical region provided by warping \mathcal{N}_{t-1} by $\mathbf{w}(\mathbf{x}; \mathbf{c}_t)$. Other parameterized image warps, and other parameterized region representations could also be used (e.g., see F.G. Meyer and P. Bouthemy, "Region-Based Tracking Using Affine Motion Models in Long Image Sequences", *CVGIP: Image Understanding*, 60(2):119-140, 1994, ~~which is incorporated herein by reference~~).

[0087] To estimate \mathbf{c}_t , the sum of the log-likelihood and the log-prior given is maximized by

$$E(\mathbf{c}_t) = L(D_t | \mathcal{A}_{t-1}, D_{t-1}, \mathbf{c}_t) + \log p(\mathbf{c}_t | \mathbf{c}_{t-1}) \quad \text{EQUATION (12)}$$

To maximize $E(\mathbf{c}_t)$ a straightforward variant of the expectation-maximization (EM) algorithm is used, as

described by A. Jepson and M. J. Black in "Mixture Models for Optical Flow Computation", In *Proc. IEEE Computer Vision and Pattern Recognition, CVPR-93*, pages 760-761, New York, June 1993, ~~which is incorporated herein by reference~~. This is an iterative, coarse-to-fine algorithm, with annealing used to control the method becoming trapped in local minima. In short, the E-step determines the ownership probabilities for the backwards warped data \hat{D}_t , as in Equation (3) above. The M-step uses these ownerships to form a linear system for the update to c_t . These components of the linear system are obtained from the motion constraints weighted by the ownership probabilities for the \mathcal{W} and \mathcal{S} processes.

[0092] In practice, to help avoid becoming stuck in local minima, it is useful to apply the EM algorithm with a coarse-to-fine strategy and deterministic annealing in fitting the motion parameters (e.g., see, for example, A. Jepson and M. J. Black, "Mixture Models for Optical Flow Computation," *Proc. IEEE Computer Vision and Pattern Recognition, CVPR-93*, pages 760-761, New York, June 1993, ~~which is incorporated herein by reference~~). The initial guess for the warp parameters is based on a constant velocity model, so the initial guess is simply equal to the estimated warp parameters from the previous frame. By way of annealing, instead of using the variances $\sigma_{s,t}^2$ and σ_w^2 in computing the ownerships and gradients of Equation (22) for the \mathcal{S} and \mathcal{W} components, the parameters σ_s and σ_w are used. At each iteration of the EM-algorithm, these values are decreased according to

$$\sigma_S \leftarrow \min(0.95\sigma_S, \hat{\sigma}_S)$$

$$\sigma_W \leftarrow \min(0.95\sigma_W, \hat{\sigma}_W)$$

EQUATION (24)

where $\hat{\sigma}_s$ and $\hat{\sigma}_w$ are the maximum likelihood variance estimates of the S component and W component phase differences, over the entire neighborhood, \mathcal{N}_t , given the motion estimate obtained in the current EM iteration. Once the variances reach a minimal value the annealing is turned off and they are allowed to fluctuate according to the current motion parameters. Moreover, as the variance of the S component decreases according to the spatial ensemble of data observations at each EM iteration, the variances used for each individual observation in computing ownerships and likelihood gradients are never allowed to be lower than the corresponding variance of $\sigma_{s,t}^2$.

[0093] Finally, once the warp parameters \mathbf{c}_t have been determined, the appearance model \mathcal{A}_{t-1} is convected (warped) forward to the current time t using the warp specified by \mathbf{c}_t . To perform this warp, a piecewise constant interpolant is used for the WSL state variables $\mathbf{m}(\mathbf{x}, t-1)$ and $\sigma_s(\mathbf{x}, t-1)$. This interpolation was expected to be too crude to use for the interpolation of the mean $\mu(\mathbf{x}, t-1)$ for the stable process, so instead the mean is interpolated using a piecewise linear model. The spatial phase gradient for this interpolation is determined from the gradient of the filter responses at the nearest pixel to the desired location \mathbf{x} on the image pyramid sampling grid (see D.J. Fleet, A.D. Jepson, and M. Jenkin, "Phase-Based Disparity Measurement," *Computer Vision and Image Understanding*, 53(2):198-210, 1991, ~~incorporated herein by reference~~).